Оптимальные градиентные методы

Мальковский Н. В.

Санкт-Петербургский академический университет

Преамбула

Общая задача минимизации

$$\begin{array}{ll} \text{минимизировать} & f(x), \\ \text{при условии} & x \in \mathcal{K}. \end{array} \tag{1}$$

Условия стационарности

Теорема

Пусть в задаче (1) *множество* $\mathcal{K}$ *выпукло, а функция* f *дифференцируема в точке* x^*, *тогда для точки минимума* x^* *выполняется условие*

$$\nabla f(x^*)^T(x - x^*) \geq 0, \quad \forall x \in \mathcal{K}.$$

Проекция шага градиентного спуска

Следствие. Если $\mathcal{K}$ – замкнутое выпуклое множество, $P_{\mathcal{K}}(x) = \text{argmin}_{y \in \mathcal{K}} ||y - x||^2$ – евклидова проекция на $\mathcal{K}$, $\alpha > 0$, то для точки минимума x^* функции f на $\mathcal{K}$ выполняется

$$x^* = P_{\mathcal{K}}(x^* - \alpha \nabla f(x^*)).$$

Док-во. Два случая:

1. $\nabla f(x^*) = 0_n$, в этом случае $x^* = x^* - \alpha \nabla f(x^*) = P_{\mathcal{D}}(x^* - \alpha \nabla f(x^*))$.
2. $\nabla f(x^*) \neq 0_n$. Пусть $z = x^* - \alpha \nabla f(x^*)$, тогда при $||y - z|| \leq \alpha ||\nabla f(x^*)||$

$$\begin{aligned} \nabla f(x^*)^T (y - x^*) &= \nabla f(x^*)^T (y - z - \alpha \nabla f(x^*)) \\ &= \nabla f(x^*)^T (y - z) - \alpha ||\nabla f(x^*)||^2 \\ &\leq ||\nabla f(x^*)|| \cdot \alpha ||\nabla f(x^*)|| - \alpha ||\nabla f(x^*)||^2 \leq 0. \end{aligned}$$

Так как единственная точка, для которой достигается равенство – x^*, а для всех точек $\mathcal{K}$ выполняется неравенство в обратную сторону, то x^* – единственная точка $\mathcal{K}$, для которой выполняется $||x^* - z|| \leq \alpha ||\nabla f(x^*)||$, таким образом $x^* = P_{\mathcal{K}}(z)$. ∎

Проективный градиентный спуск

Простая адаптация метода градиентного спуска для задач на ограниченных множествах:

$$x_{k+1} = P_{\mathcal{K}}(x_k - \alpha_k \nabla f(x_k)). \tag{2}$$

Такой метод отыскания точки минимума называется методом проекции градиента или проективным градиентным спуском. В случае, когда $\mathcal{K}$ – замкнутое выпуклое множество, основные свойства градиентного спуска сохраняются и при использовании проективного метода.

Свойства оператора проекции

Пусть $\mathcal{K}$ – замкнутое выпуклое множество, $P_{\mathcal{K}}$ – евклидова проекция на $\mathcal{K}$, тогда

1. $(P_{\mathcal{K}}(x) - x)^T(y - P_{\mathcal{K}}(x)) \geq 0 \;\; y \in \mathcal{K}$.
2. $||P_{\mathcal{K}}(x) - P_{\mathcal{K}}(y)|| \leq ||x - y||$.

Док-во. По определению $P_{\mathcal{K}}(x) = \operatorname{argmin}_{y \in \mathcal{K}} ||y - x||^2$. Из условий стационарности

$$\nabla[||y - x||^2]^T_{y = P_{\mathcal{K}}(x)}(z - P_{\mathcal{K}}(x)) \geq 0 \;\; \forall z \in \mathcal{K}.$$

Первый пункт напрямую следует из этого неравенства и $[||y - x||^2]_{y = P_{\mathcal{K}}(x)} = P_{\mathcal{K}}(x) - x$.

Свойства оператора проекции

Используя свойство 1 получаем

$$(P_{\mathcal{K}}(x) - x)^T (P_{\mathcal{K}}(y) - P_{\mathcal{K}}(x)) \geq 0$$
$$(P_{\mathcal{K}}(y) - y)^T (P_{\mathcal{K}}(x) - P_{\mathcal{K}}(y)) \geq 0$$

Вычитая первое из второго получаем

$$(P_{\mathcal{K}}(y) - P_{\mathcal{K}}(x) - (y - x))^T (P_{\mathcal{K}}(x) - P_{\mathcal{K}}(y)) \geq 0,$$

$$(x - y)^T (P_{\mathcal{K}}(x) - P_{\mathcal{K}}(y)) \geq ||P_{\mathcal{K}}(x) - P_{\mathcal{K}}(y)||^2.$$

Применяя неравенство Коши-Буняковского-Шварца получаем свойство 2. ∎

Анализ сходимости

Теорема

Пусть $f : \mathcal{D} \subset \mathbb{R}^n \to \mathbb{R}$, $\mathcal{K} \subset \mathcal{D}$, f *дважды дифференцируема при этом* $mI \preceq \nabla^2 f(\cdot) \preceq MI$ *на* $\mathcal{K}$, $0 < \alpha \leq \frac{2}{M+m}$, *тогда для последовательности* x_k, *генерируемой по правилу* (2), *выполняется*

$$||x_k - x^*|| \leq (1 - \alpha M)^k \, ||x_0 - x^*||$$

Док-во.

$$\begin{aligned}
||x_{k+1} - x^*|| &= \|P_{\mathcal{K}}(x_k - \alpha\nabla f(x_k)) - P_{\mathcal{K}}(x^* - \alpha\nabla f(x^*))\| \\
&\leq \|x_k - x^* - \alpha(\nabla f(x_k) - \nabla f(x^*)\| \\
&= \left\| x_k - x^* - \alpha \int_0^1 \nabla^2 f(x^* + t(x_k - x^*))(x_k - x^*) dt \right\| \\
&\leq ||x_k - x^*|| \cdot \left\| I - \alpha \underbrace{\int_0^1 \nabla^2 f(x^* + t(x_k - x^*)) dt}_{A_k} \right\|
\end{aligned}$$

По условию $mI \preceq A_k \preceq MI$, таким образом если сходимость имеет место при $\alpha < \frac{2}{M}$. Если при этом $\alpha < \frac{2}{M+m}$, то спектр матрицы $I - \alpha A_k$ лежит на отрезке $[1 - \alpha M, 1 - \alpha m$, при этом $|1 - \alpha m| > |1 - \alpha M|$, что дает

$$||x_{k+1} - x^*|| \leq (1 - \alpha M)||x_k - x^*|| \quad \blacksquare$$

Оптимальная схема проективного градиентного спуска

По аналогии с оптимальными схемами градиентного спуска можно построить схему для проективного случая с линейной сходимостью с показателем $\frac{\sqrt{M}-\sqrt{m}}{\sqrt{M}+\sqrt{m}}$:

Инициализация Выбрать начальное приближение $x_0 \in \mathcal{K}$, $\alpha_0 \in (0,1)$. Взять $y_0 = x_0$.

Итерация $k \le 0$

1. Вычислить $\nabla f(y_k)$, взять $x_{k+1} = P_{\mathcal{K}}(y_k - \frac{1}{M}\nabla f(y_k))$.
2. Вычислить α_{k+1} из уравнения

$$\alpha_{k+1}^2 = (1-\alpha_{k+1})\alpha_k^2 + \frac{m}{M}\alpha_{k+1}$$

3. Взять

$$\beta_k = \frac{\alpha_k(1-\alpha_k)}{\alpha_k^2 + \alpha_{k+1}}, \quad y_{k+1} = x_{k+1} + \beta_k(x_{k+1} - x_k).$$

Барьерные функции

Определение

<u>Барьерной функцией</u> замкнутого множества $\mathcal{K}$ называется любая функция $f : \operatorname{Int}\mathcal{K} \to \mathbb{R}$, удовлетворяющая

$$f(x) \xrightarrow{x \to \mathcal{K} \setminus \operatorname{Int}\mathcal{K}} +\infty$$

Замечание Барьерная функция множества является приближением функции-индикатора этого множества.

Центральный путь

Рассмотрим задачу (1). Пусть F – барьер $\mathcal{K}$. Посмотрим на вспомогательную задачу оптимизации с параметром t

$$\text{минимизировать} \quad f(x) + \frac{1}{t}F(x) \tag{3}$$

Определение

Центральным путем задачи (1) *и барьера* F *называется кривая* $\varphi : [0, +\infty) \to \operatorname{Int} \mathcal{K}$

$$\varphi(t) = \operatorname{argmin}_{x \in \operatorname{Int} \mathcal{K}} f(x) + \frac{1}{t}F(x)$$

Замечание. Чем больше t, тем меньше второе слагаемое влияет на точку минимума (3).

Центральный путь

Теорема

Пусть φ *– центральный путь задачи* (1) *с барьером* F, $F(x) \geq F^* > -\infty$, x^* *– решение* (1), *тогда*

$$f(\varphi(t)) \xrightarrow{t \to +\infty} f(x^*).$$

Док-во. Пусть $x \in \int \mathcal{K}$, тогда

$$\overline{\lim_{t\to\infty}} f(\varphi(t)) \leq \lim_{t\to\infty} \left[f(x) + \frac{1}{t} F(x) \right] = f(x).$$

Переходя к супремуму и пользуясь непрерывностью f получаем $\overline{\lim}_{t\to\infty} f(\varphi(t)) \leq f(x^*)$.
С другой стороны

$$f(\varphi(t)) = \min_{x \in \mathcal{K}} \left\{ f(x) + \frac{1}{t} F(x) \right\} \geq \min_{x \in \mathcal{K}} \left\{ f(x) + \frac{1}{t} F^* \right\} = f(x^*) + \frac{1}{t} F^*. \quad \blacksquare$$

Стандартные барьеры

Пусть множество $\mathcal{K}$ задано неравенством $g(x) \leq 0_m$. Барьерными функциями этого множества могут служить следующие:

- Степенной барьер: $p > 1$

$$F(x) = \sum_{i=1}^{m} \frac{1}{(-g_i(x))^p}$$

- Логарифмический барьер:

$$F(x) = \sum_{i=1}^{m} -\ln(-g_i(x)^p)$$

- Экспоненциальный барьер:

$$F(x) = \sum_{i=1}^{m} \exp\left(\frac{1}{-g_i(x)^p}\right)$$

Барьерный метод

Общая схема барьерного метода для задачи (1) имеет следующий вид

1. Выбрать последвательность $t_0 < t_1 < \ldots$ такую, что $t_k \to \infty$.
2. Вычислить $x_k = \varphi(t_k)$.

Методы внутренней точки

На практике вычисление точного значения $\varphi(t)$ невозможно (лишь с заданной точностью). Обычно вместо общей схемы барьерного метода попеременно меняется $t_k \to t_{k+1}$ и делается шаг в сторону оптимума $\varphi(t_{k+1})$:

Итерация $k = 1, 2, \ldots$:

- $x_{k+1} = update(x_k, f, g, F)$

Замечание. В методах внутренней точки в качестве *update* обычно используется шаг метода Ньютона для $f(x_k) + \frac{1}{t_k} F(x_k)$. Важно отметить, что для логарифмического барьера шаг Ньютона гарантирует, что $x_{k+1} \in \mathcal{K}$.